PREFACE

SAC '94 is the first of an annual series of workshops on Selected Areas in Cryptography. The topics chosen for this first SAC Workshop are:

- Design and Analysis of Secure Private-Key Block Ciphers,
- Formal Methods for Cryptographic Protocols, and
- Related Topics.

The intent of the Workshop is to bring together researchers in cryptography to present new work on a few areas of current interest. An effort will be made to provide an opportunity for in-depth discussion in a relaxed atmosphere.

SAC '94 is being held at Queen's University in Kingston, with plans to hold SAC '95 at Carleton University in Ottawa, May 1995. It is expected that the site will alternate between these two locations. We hope that the SAC Workshops will complement the other major conferences in cryptography which cover the general field. An extended abstract of the papers presented at the SAC Workshops will be printed and made available to attendees as a Workshop Record. A limited number of copies of the Record will be sold to those who request it.

We wish to thank the Department of Electrical and Computer Engineering at Queen's University and the Telecommunication Research Institute of Ontario (TRIO) for their administrative support. We also wish to thank Tracey Livingstone of the TRIO Office at Queen's University for her help with producing the Workshop Record and help with registration.

On behalf of the Organizing Committee consisting of Carlisle Adams, Henk Meijer, Paul van Oorschot and myself, I would like to welcome you to SAC '94, Queen's University and Kingston.

Stafford Tavares
Queen's University

WORKSHOP ON SELECTED AREAS IN CRYPTOGRAPHY (SAC '94)

May 5 & 6, 1994

PROGRAM

(Sessions will take place in Walter Light Hall, Queen's University, Rooms 210 and 212.)

Thursday, May 5

10:30 Welcoming Remarks and Introduction of the Invited Speaker
Stafford Tavares

10:45 Invited Speaker: **Michael Wiener,** ***Bell Northern Research***
Efficient DES Key Search

SECRET-KEY CRYPTOSYSTEMS I 1
Chair: Carlisle Adams

11:35 **L. O'Connor,** ***Queensland University of Technology***
Designing Product Ciphers using Markov Chains

12:00 **J. Seberry, X.-M. Zhang, and Y. Zheng,** ***The University of Wollongong***
Nonlinearity Characteristics of Quadratic Substitution Boxes

12:25 - 2:00 Lunch

CRYPTOGRAPHIC PROPERTIES OF BOOLEAN FUNCTIONS 30
Chair: Howard Heys

2:00 **D. Stinson,** ***University of Nebraska***
Recent Results on Resilient Functions

2:25 **M. Zhang, S.E. Tavares, and L.L. Campbell,** ***Queen's University***
Information Leakage of Boolean Functions as a Measure of Cryptographic Strength

2:50 **J. Seberry, X.-M. Zhang, and Y. Zheng,** ***The University of Wollongong***
How to Better the SAC

3:15 - 3:45 Coffee Break

SECRET-KEY CRYPTOSYSTEMS II 59
Chair: Jennifer Seberry

3:45 **E.P. Dawson, L.J. O'Connor, and H.M. Gustafson,** ***Queensland University of Technology***
Linearity in Block Ciphers

4:10 **K. Kim, S. Lee, S. Park, and D. Lee,** ***ETRI, Korea***
DES Can be Immune to Linear Cryptanalysis

4:35 **J. Pieprzyk, C. Charnes, and J. Seberry,** ***The University of Wollongong***
Linear Approximation Versus Nonlinearity

5:00 End of Thursday's Program

7:00 SOCIAL at The University Club

8:00 DINNER at The University Club

Friday, May 6

8:30 Morning Coffee and Snacks

9:00 Announcements and Introduction of the Invited Speaker
Paul van Oorschot

9:10 Invited Speaker: **Richard Kemmerer,** ***University of California at Santa Barbara***
Using Formal Methods to Analyze Encryption Protocols

ANALYSIS OF CRYPTOGRAPHIC PROTOCOLS I 94
Chair: Henk Meijer

10:00 **P. Syverson,** ***Naval Research Laboratory, Washington, D.C.***
A Taxonomy of Replay Attacks

10:25 **W. Mao and C. Boyd,** ***University of Manchester***
Classification of Cryptographic Techniques in Authentication Protocols

10:50 - 11:20 Coffee Break

ANALYSIS OF CRYPTOGRAPHIC PROTOCOLS II 107
Chair: Paul Syverson

11:20 **D. Stal, S.E. Tavares, and H. Meijer,** ***Queen's University***
Backward State Analysis of Cryptographic Protocols Using Coloured Petri Nets

11:45 **P. Syverson and P. van Oorschot,** ***Naval Research Laboratory, Carleton University, and Bell Northern Research***
On Unifying Some Cryptographic Protocol Logics

12:10 - 2:00 Lunch

SECRET-KEY CRYPTOSYSTEMS III 120
Chair: Ed Dawson

2:00 **J. Benaloh,** ***Clarkson University***
Dense Probabilistic Encryption

2:25 **C.M. Adams,** ***Bell Northern Research***
Simple and Effective Key Scheduling for Symmetric Ciphers

2:50 **H.M. Heys and S.E. Tavares,** ***Queen's University***
Key Clustering in Substitution-Permutation Network Cryptosystems

3:15 - 3:30 Coffee Break

Chair: Josh Benaloh

3:30 **W. Millan, E.P. Dawson, and L.J. O'Connor,** ***Queensland University of Technology***
Fast Attacks on Tree-Structured Ciphers

3:55 **J.-Y. Chouinard and G. Ferland,** ***University of Ottawa***
Cryptographic Degradation of DES in Block and Stream Cipher Modes in a Digital Mobile Communication Link

4:20 Closing Remarks

4:30 End of Workshop

SECRET-KEY CRYPTOSYSTEMS I

Efficient DES Key Search

Michael J. Wiener

Bell-Northern Research, P.O. Box 3511 Station C, Ottawa, Ontario, K1Y 4H7, Canada

1994 March 31

Abstract. Despite recent improvements in analytic techniques for attacking the Data Encryption Standard (DES), exhaustive key search remains the most practical and efficient attack. Key search is becoming alarmingly practical. We show how to build an exhaustive DES key search machine for $1 million that can perform a known-plaintext attack in 3.5 hours on average. This machine contains 57600 special-purpose DES key search chips. The chip and the rest of the machine have been designed in detail for the purpose of assessing the resistance of DES to an exhaustive attack; we have no plans to build the machine. This design is based on mature technology to avoid making guesses about future capabilities. With this approach, DES keys can be found one to two orders of magnitude faster than other recently proposed designs.

The basic machine design can be adapted to attack the standard DES modes of operation for a small penalty in running time. A $1 million machine would take 8 hours on average to find a key used in 1-bit CFB mode and 4 hours on average for any of ECB, CBC, 64-bit OFB, 64-bit CFB, or 8-bit CFB mode.

In the past, a concern about key search machines was that they would break down too frequently to produce any useful results. This is not a problem with current technology. The expected failure rate of the DES key search machine described here is one failure for every 270 keys found.

If it ever was true that attacking DES was only within the reach of large governments, it is clearly no longer true. In light of this work, it would be prudent in many applications to use DES in a triple-encryption mode.

Designing product ciphers using Markov Chains

Luke O'Connor*
Distributed Systems Technology Centre
and
Information Security Research Center, QUT

Abstract

In this paper we consider the design of product ciphers based on Markov chains. We examine two particular chains which are related to the differential and linear cryptanalysis attacks. Both of these chains approach the uniform distribution which indicates that appropriately designed ciphers are secure against these attacks. The maximum deviation from the uniform distribution can be used as guide for the number of rounds the cipher should iterate.

1 Introduction

Consider an R-round product cipher with round function F, such that the cipher operates by iterating F R times. A system designer is likely to ask the question'when are $R+1$ rounds better than R rounds?' Let us assume that there is some means of measuring the 'goodness' of ciphertext produced from an R-round block cipher. There are several established criteria that could be used here. For a given plaintext bit p_i we may consider the number of ciphertext bits that depend on p_i, which will be an integer in the range $[0, n]$ for n-bit ciphers. Also, each ciphertext bit $c_i = f_i(P)$ can be expressed as function of the plaintext P for a fixed key K, and we may consider the nonlinearity of f_i which is an integer in the range $[0, 2^{n-1} - 2^{\lceil \frac{n}{2} \rceil - 1}]$. It is then clear that for each measure M with N possible outcomes $q_1, q_2, \ldots, q_N$, we can associate a distribution $\Pi = (\pi_1, \pi_2, \ldots, \pi_N)$ such that the measure M after round R is equal to q_i with probability π_i. In the cases of variable dependency and nonlinearity these distributions should be very skewed in that

*The work reported in this paper has been funded in part by the Cooperative Research Centres program through the Department of the Prime Minister and Cabinet of Australia. Mailing address: Information Security Research Center, Queensland University of Technology, GPO Box 2434, Brisbane, Qld. 4001, Australia. Email: oconnor@fitmail.fit.qut.edu.au.

strong ciphers are likely to take on the larger values with high probability. That is, with high probability strong cipher should depend on most variables, and be a large distance from the set of linear functions. A guide to determining the value of R would be to iterate until the probability distribution Π has a form which suggests the measure is close to an optimal value.

In some cases it is better that the measure have a more uniform rather than skewed distribution, suggesting that it is difficult to distinguish between possible outcomes of the measure. Consider encrypting two plaintexts P and P' which only differ in one bit. If the cipher is acting like a good mixing transformation should, the respective ciphertexts should appear to be uncorrelated after a (possibly large) number of rounds. This phenomenon is called the *avalanche effect* and refers to the propagation of small changes in the cipher (here 1 bit) leading to large unpredictable changes in the ciphertext. More formally, consider the set of differences $\Delta C_r = C_r + C'_r$ generated when $C_r(C'_r)$ is the encryption of $P(P')$ after r rounds. If there are N possible differences that ΔC_r could assume then a good mixing transformation would distribute ΔC_r with approximately equal probability amongst the N differences. Then for large r we would hope to have that Π is close to the uniform distribution, or $\Pi \approx (1/N, 1/N, \ldots, 1/N)$.

Note that the entropy of Π, denoted $H(\Pi)$, is maximized when Π is the uniform distribution. There is elegant result from information theory which states that if $\mathbf{P}$ is a doubly stochastic matrix, then $H(\mathbf{P} \cdot \Pi) \geq H(\Pi)$ with equality only when Π and $\mathbf{P} \cdot \Pi$ are rearrangements of each other. This suggests that the designer should attempt to associate a doubly stochastic matrix $\mathbf{P}$ with the round function F, such that after r rounds $\Pi = \mathbf{P}^r \cdot \Pi_0 = \mathbf{P}(\mathbf{P}^{r-1} \cdot \Pi_0)$, where Π_0 is the initial distribution of the measure. With a few additional constraints on $\mathbf{P}$ (discussed later), the measure distribution Π will tend to the uniform distribution as the entropy is increasing after each additional round. Since $\Pi = \mathbf{P}^r \cdot \Pi_0$ it follows that the round function F must form a Markov chain with respect to the measure M. Intuitively, this is not wholly unexpected since the operation of the round function is typically fixed, taking only the current ciphertext and subkey as parameters.

While it is a simple matter to state this Markov design method, the reader will remark that it is much harder to apply. Surprisingly, there are many product ciphers that will have Markov chains with exactly the properties stated when considering both differential and linear cryptanalysis [3, 6, 7]. Recall that our original problem was to determine when it is worthwhile to add one more round in a product cipher. Well, if we know that Π is tending to the uniform distribution then we can fix a deviation ϵ from this distribution and iterate until

$$|(1/N, 1/N, \ldots, 1/N) - \mathbf{P}^r \cdot \Pi_0| \leq \epsilon. \tag{1}$$

The deviation expressed in (1) is concerned with the convergence of the chain to its

limiting distribution, a problem which has been thoroughly studied (see Bhat [2] for the classical approach and Vizarani [14] for recent results). In the remainder of the paper we will discuss particular chains that can be defined for differential and linear cryptanalysis and consider the class of ciphers for which these chains apply and their convergence properties.

2 A chain for differential cryptanalysis

We begin with considering a Markov chain for differential cryptanalysis. Much of this discussion will apply to the chain for linear cryptanalysis. Assuming that there are N possible states in the chain, the one-step (one round) transition probabilities can be described in a $N \times N$ matrix $\mathbf{P} = [P_{ij}]$, $1 \leq i, j \leq N$,

$$\mathbf{P} = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1N} \\ P_{21} & P_{22} & \cdots & P_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ P_{N1} & P_{N2} & \cdots & P_{NN} \end{bmatrix}. \tag{2}$$

For differential cryptanalysis, the $\mathbf{P}$ matrix is directly obtained from the XOR table for the round function F. For an n-bit round function F, for each input difference $\Delta P = i$ and output difference $\Delta C = j$, $1 \leq i, j \leq 2^n - 1$, let P_{ij} be defined as

$$P_{ij} = 2^{-n} \cdot \sum_{\substack{X, X' \in Z_2^n \\ \Delta P = X + X'}} [F(X) + F(X') = \Delta C] \tag{3}$$

where $[\cdot]$ is a boolean predicate evaluating to 0 or 1. Here $N = 2^n - 1$ since we do not consider the degenerate cases where $i = 0$ or $j = 0$ (that is, the states of the chain correspond to the $2^n - 1$ nonzero n-bit vectors). Note that for DES this means that $\mathbf{P}$ has dimensions $(2^{64} - 1) \times (2^{64} - 1)$ which is very large indeed. If $P_{ij} > 0$ we will write $\Delta P \rightarrow \Delta C$, meaning that in one round it is possible for an input difference of ΔP to lead to an output difference of ΔC. Consider the probability of the event where a plaintext difference of ΔP leads to a ciphertext difference of ΔC_r after r rounds, described as

$$\Pr(\Delta P \rightarrow \Delta C_r) = \Pr(\Delta P \rightarrow \Delta C_1 \rightarrow \Delta C_2 \rightarrow \cdots \rightarrow C_{r-1} \rightarrow C_r). \tag{4}$$

We are unconcerned about the actual values taken on by the intermediate differences $\Delta C_1, \ldots, \Delta C_{r-1}$, only that they provide a valid state transition path from ΔP to ΔC_r. We call the pair $(\Delta P, \Delta C_r)$ an *r-round differential*. It is clear that the RHS of (4) is

stochastic but it is a jump in logic to see that it is in fact Markovian. The Markovian property states that when $\Delta P = \Delta C_0$

$$\Pr(\Delta P \rightarrow \Delta C_r) \;=\; \sum_{\Delta C_1,\ldots,\Delta C_{r-1}} \prod_{k=1}^{r} \Pr(\Delta C_{i-1} \rightarrow \Delta C_i). \tag{5}$$

That is, the probability of the chain of differences given in (4) is equal to the product of probabilities for the single round differences $\Delta C_{i-1} \rightarrow \Delta C_i$. This would be true if we could somehow arrange for the rounds to operate 'independently', so that the actual ciphertext does not cause the events $\Pr(\Delta C_{i-2} \rightarrow \Delta C_{i-1})$ and $\Pr(\Delta C_{i-1} \rightarrow \Delta C_i)$ to be dependent. For ciphers such as DES, FEAL and LOKI, the Markov property is proven by considering the subkeys XORed to the ciphertext at each round. If these subkeys $K_1, K_2, \ldots, K_r$ are assumed to be independent then the particular input to a given round becomes random, which means that the ciphertext pair defining the current state is random across the 2^n pairs that could define the state, and the process then becomes Markovian. Of course it is possible to construct round functions for which *independent subkeys will not induce the Markovian property* but conveniently, all DES-like ciphers of interest have this property. It then follows from (5) that

$$\Pr(\Delta P \rightarrow \Delta C_r) \;=\; P_{ij}^{(r)}. \tag{6}$$

where $\mathbf{P}^{(r)} = \mathbf{P}^r = [P_{ij}^{(r)}]$. The task of showing that $\Pr(\Delta P \rightarrow \Delta C_r)$ is tending to some small value as a function of r is now simplified to the study of the $\mathbf{P}$ matrix, which we may attack using the full theory of finite Markov chains. The class of ergodic chains (defined below) is of particular intertest to us since the asymptotic behaviour of $\mathbf{P}^{(r)}$ has been determined.

Theorem 2.1 If $\mathbf{P}$ is ergodic then there exists a unique distribution $\Pi = (\pi_1, \pi_2, \ldots, \pi_N)$ such that

$$\pi_j \;=\; \lim_{r\rightarrow\infty} P_{ij}^{(r)}. \tag{7}$$

The distribution Π is said to the *limiting distribution* for $\mathbf{P}$. □

A limiting distribution implies that, regardless of the initial state of the chain, the probability that the chain ends up in state j at time r is tending towards π_j for large n. Observe that $\sum_j P_{ij} = 1$ by definition, and if in addition $\sum_i P_{ij} = 1$, then $\mathbf{P}$ is said to be *doubly stochastic*.

Lemma 2.1 Let $\mathbf{P}$ be an $N \times N$ doubly stochastic matrix modeling an ergodic process. Then the uniform distribution $\Pi = (1/N, 1/N, \cdots, 1/N)$ is the stationary distribution for $\mathbf{P}$. □

Since the round function F is bijective, it is easily shown that $\mathbf{P}$ is doubly stochastic. Then if $\mathbf{P}$ were ergodic, it could be shown that all differentials $(\Delta P, \Delta C_r)$ are tending to be equally likely when enough rounds are used. Such ciphers are rendered immune to differential cryptanalysis since the event $\Delta P \rightarrow \Delta C_r$ is indistinguishable from the event of ΔP leading to an arbitrary difference after r rounds. In the next section we will show that $\mathbf{P}$ is ergodic for almost all round functions F.

2.1 Demonstrating the Ergodic property

Some standard definitions must be recalled at this point. State i communicates with state j if $P_{ij}^r > 0$ for some r, denoted by $i \leftrightarrow j$. It can be verified that '$\leftrightarrow$' is an equivalence relation, and if $\mathbf{P}$ has only one equivalence class (all states communicate), then $\mathbf{P}$ is said to be *irreducible*. State i has period d_i if $P_{ii}^r = 0$ whenever r is not divisible by d_i; also, i is said to aperiodic if $d_i = 1$. The period of $\mathbf{P}$ is defined as $d = \gcd(d_1, d_2, \cdots, d_N)$, and $\mathbf{P}$ is said to be aperiodic if $d = 1$. Finite, irreducible, aperiodic chains are called *ergodic*. Since $\mathbf{P}$ is clearly finite, to prove ergodicity, we must demonstrate that the chain is both aperiodic and irreducible. Conveniently, the aperiodic property is easily verified for all $\mathbf{P}$ from Minc's observation that all doubly stochastic matrices have a nonzero entry on the diagonal [8].

Irreducibility is a much harder property to demonstrate in general. Since the size of $\mathbf{P}$ is expected to be close to $2^{128} = 2^{64} \times 2^{64}$ we must use some probabilistic argument to show that $\mathbf{P}$ is irreducible. Luckily, we may pass to random graph theory to do exactly that. As has been observed by many authors, $\mathbf{P}$ can be considered as the adjacency matrix for a directed graph $\mathbf{G} = (V, E)$, where $V = \{v_1, v_2, \cdots, v_N\}$ and there is a directed edge from v_i to v_j if and only if $P_{ij} > 0$. We will call $\mathbf{G}$ the *underlying graph* of $\mathbf{P}$.

Proposition 2.1 The directed graph $\mathbf{G}$ is strongly connected if for all v_i, v_j, there is a directed path from vertex v_i to vertex v_j. Then the matrix $\mathbf{P}$ is irreducible if and only if $\mathbf{G}$ is strongly connected. □

One way to argue that $\mathbf{G}$ is strongly connected would be to show that $\mathbf{G}$ has a sufficiently large number of edges so as to ensure that directed paths exist between all vertex pairs with high probability. More briefly, strong connectivity is almost certain when the number of edges is large. The following result is due to Palásti, and is also reported by Bollobás [4].

Theorem 2.2 (Palásti [12]) Let $m = N\{\log N + c + o(1)\}$ for some real c, and let G be selected uniformly from all N-vertex graphs with m edges. Then for large N,

$$\Pr(G \text{ is strongly connected}) \rightarrow e^{-2e^{-c}} \tag{8}$$

where e is the base of the natural logarithm. □

We will assume that that the entries of $\mathbf{P}$ are distributed approximately randomly with respect to being zero or nonzero, as all computational results suggest. We then claim that if the number of edges in $\mathbf{G}$ dominates $N \log N$ then $\mathbf{G}$ is strongly connected with high probability.

Theorem 2.3 (Oconnor [10]) Let λ^* be the largest entry in the XOR table for a bijective $F : Z_2^n \to Z_2^n$. Then if F is selected uniformly, $\lim_{n\to\infty} \frac{\mathbf{E}[\lambda^*]}{2n} \leq 1$. □

Corollary 2.1 If F is selected uniformly, then $\Pr(\mathbf{P} \text{ is ergodic}) \to 1$.

Proof. Since $\mathbf{E}[\lambda^*] = \sum_i i \cdot \Pr(\lambda^* = i)$, it follows from Theorem 2.3 that $\Pr(\lambda^* \leq 2n) \to 1$ for large n as

$$\lim_{n\to\infty} \frac{\mathbf{E}[\lambda^*]}{2n} \leq 1 \implies \sum_{i>2n} i \cdot \Pr(\lambda^* = i) \to 0 \implies \sum_{i>2n} \Pr(\lambda^* = i) \to 0.$$

Then with probability tending to 1 the XOR table will have at least $\frac{2^{2n}}{2n} = \Theta(N^2/\log N)$ nonzero entries which asymptotically dominates the $N \log N$ bound from Theorem 2.2. □

Let us review our progress thus far. We initially showed that defining a measure M which induced a Markov chain with a doubly stochastic transition matrix $\mathbf{P}$ on the round function F was a good criterion since the distribution Π of the measure M is tending to the uniform distribution. The specific measure we considered in this section was the distribution of differences defined by the Δ operator, defined naturally from the XOR table of the round function. To demonstrate that the chain converges to the uniform distribution we had to prove that the chain was ergodic. Of the defining properties for ergodicity, only irreducibility is difficult to verify, and to do this we have used results from random graph theory to suggest the strong connectivity of $\mathbf{P}$ for almost all round functions F. Most of the analytical work is in the counting argument used to suggest strong connectivity.

We will now take similar steps to define an ergodic chain with respect to linear cryptanalysis. Again, most of the work is also in the counting argument used to suggest strong connectivity, but we will not present those details here. Our main observation is that the probability q^* of a successful linear cryptanalysis can be cast in terms of correlation coefficients $c(\cdot,\cdot)$, and there is a Markov chain $\mathbf{P}$ which upper bounds the approximation expressed in these coefficients.

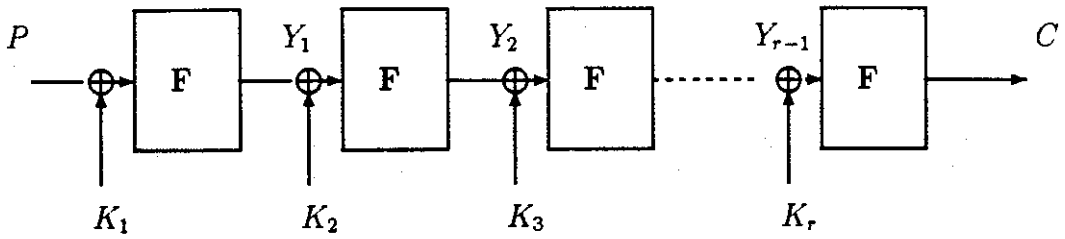

Figure 1: A general product block cipher.

3 A chain for Linear Cryptanalysis

Linear cryptanalysis was introduced recently by Matsui [6, 7], and is currently the best known non-exhaustive attack on the DES. The basis of the attack is finding linear relationships between certain bits of the plaintext P, ciphertext C and key K, expressed as $\sum P + \sum C = \sum K$. More formally, the attack derives an approximation of the form

$$\sum_{i=1}^{n} a_{1i}p_i \; + \; \sum_{i=1}^{n} a_{2i}c_i \;\; = \;\; \sum_{i=1}^{m} a_{3i}k_i \qquad (\mathrm{mod}\, 2) \tag{9}$$

where $a_{ij} \in \{0, 1\}$ and $P = p_1, \ldots, p_n$, $C = c_1, \ldots, c_n$ and $K = k_1, \ldots k_m$. Matsui has shown that $\sum K$ can be determined accurately using the maximum likelihood method if a sufficient amount N_L of known ciphertext is available. In particular, if the approximation in (9) is correct with probability q^*, the attack is expected to be successful 98% of the time when $N_L \approx |q^* - 1/2|^{-2}$. The result of the attack is the knowledge of one bit of information concerning the key, namely the value of $\sum K$.

We will now consider how the approximation in (9) is found. Our discussion will be with respect to the cipher shown in Figure 1, but applies to more practical ciphers such as DES. A linearization τ of a mapping, such as an S-box, is an approximate relation between a sum of its inputs and a sum of its outputs. An approximation similar to (9) for an r round cipher is found by determining r linearizations $\tau_1, \tau_2, \cdots, \tau_r$ of the F function, with the property that $\sum_i \tau_i \pmod 2$ only involves plaintext, ciphertext and key bits as unknowns. That is, all terms involving input or output to internal rounds of the cipher that cannot be represented as plaintext or ciphertext cancel. Our main observation is that this cancellation property allows the probability of $\sum_i \tau_i$ to be upper bounded by an appropriately defined Markov chain.

Consider the following linearizations as depicted in Figure 1. Let $Y_i = y_{i,1}, y_{i,2}, \ldots, y_{i,n}$ be a binary n-vector representing a linear function of n binary variables, for $0 \leq i \leq r$. Also let the intermediate ciphertext at round i be denoted as $C_i = c_{i,1}, c_{i,2}, \ldots, c_{i,n}$, with

the subkey denoted as $K_i = k_{i,1}, k_{i,2}, \ldots, k_{i,n}$. Let the linearization τ_i at round i, between Y_{i-1} and Y_i, $1 \leq i \leq r$, be

$$\sum_{j=1}^{n} y_{i-1,j}(c_{i-1,j} \oplus k_{i,j}) \;=\; \sum_{j=1}^{n} y_{i,j}c_{i,j}, \tag{10}$$

which is true with some probability q_i. Note that Y_{i-1} denotes the sum of the input variables used in the approximation at round i, and also the sum of the output variables used in the approximation at round $i-1$, for $i > 1$. When these r approximations are added modulo 2 the two terms involving $Y_i 1 \leq i \leq r-1$, will cancel, leaving an approximation in terms of a sum of plaintext bits $Y_0 = \sum P$, a sum of the ciphertext bits $Y_r = \sum C$, and a sum of key bits $\sum K$, which is the same form as in (9). Let this approximation be correct with probability q^*.

When the keys bits are assumed to be independent, the approximations in (10) for different i are independent, due to the way the subkey and the current ciphertext are combined at each round. In this case, Gallager [5] has shown that

$$q^* \;=\; \frac{1}{2} + 2^{r-1} \cdot \prod_{i=1}^{r} \left(q_i - \frac{1}{2}\right) \tag{11}$$

referred to as the Piling-Up lemma by Matsui. Note that both differential and linear cryptanalysis derive r-round approximations by combining r 1-round approximations. And in both attacks, the probability of the r-round approximation is derived from a product of the 1-round approximations, made possible by the independent subkeys which give the approximations a Markovian character.

It is more convenient to reformulate the probability q^* in terms of correlation coefficients. If f and g are two n-bit boolean functions then the correlation coefficient, or simply, the correlation of f and g, denoted by $c(f,g)$, is defined as

$$c(f,g) \;=\; 2^{-n} \cdot \sum_{X} (-1)^{f(X)}(-1)^{g(X)} \;=\; \Pr(f = g) - \Pr(f \neq g).$$

But if for each i we view Y_{i-1} and Y_i as boolean functions $\sum_{j=1}^{n} y_{i-1,j}c_{i-1,j}$ and $\sum_{j=1}^{n} y_{ij}c_{i,j}$, respectively, then $q_i - \frac{1}{2} = c(Y_{i-1}, Y_i)/2$ and q^* can then be written as

$$q^* = \frac{1}{2} + \frac{1}{2} \cdot \prod_{i=1}^{r} c(Y_{i-1}, Y_i). \tag{12}$$

Recall that the complexity of linear cryptanalysis is $N_L \approx |q^* - 1/2|^{-2}$, and hence a bound on the deviation of q^* from one half would be important in determining the complexity of the attack. From (12) we see that this deviation from one half depends on the product $\prod_{i=1}^{r} c(Y_{i-1}, Y_i)$. Our main result is to show that there is a Markov chain $\mathbf{P}$ such that $\prod_{i=1}^{r} c(Y_{i-1}, Y_i) \leq P_{ij}^{(r)}$ for some states i, j, such that the limiting distribution of $\mathbf{P}$ is the uniform distribution.

3.1 The correlation matrix

The round function $F : Z_2^n \to Z_2^n$ is bijective, and let $f_i : Z_2^n \to Z_2$ be the boolean function that describes the ith bit of F, $1 \leq i \leq n$. Let $X = x_1 x_2 \cdots x_n$ denote the vector of input variables. For $N = 2^n - 1$, define an $N \times N$ matrix $\mathbf{P} = [P_{ij}], 1 \leq i, j \leq N$, where $i = (i_1 i_2 \cdots i_n)_2, j = (j_1 j_2 \cdots j_n)_2$ and

$$P_{ij} = c^2\left(\sum_k i_k x_k, \sum_k j_k f_k\right). \tag{13}$$

That is, P_{ij} is the square of the correlation between a particular linear combination of the x_i and a linear combination of the f_i. Note that the chain excludes the trivial values of $i = j = 0$. We will call $\mathbf{P}$ the correlation matrix. The sum of a given column corresponds to the total correlation between $f = \sum_k j_k f_k$ and the set of all nontrivial linear functions. Since f is balanced this is also the total correlation to all the linear functions which is known to be one for any boolean function [13]. On the other hand, since F is bijective, input variables can be expressed as a bijective function, F^{-1} of the output variables. By the same argument one then obtains that each row of $\mathbf{P}$ also sums to one. Hence we have

Lemma 3.1 For a bijective function F, the correlation matrix $\mathbf{P}$ is doubly stochastic. □

Let $Y = (Y_0, Y_1, \cdots, Y_r)$ denote an r-round approximation of a block cipher. Let $\mathbf{P}^{(r)} = [P_{ij}^{(r)}]$ be the rth power of $\mathbf{P}$. By definition of the chain we have that

$$P_{ij}^{(r)} = \sum_{\substack{\sum i_k P_k = Y_0 \\ \sum j_k c_k = Y_r}} \prod_{t=1}^{r} c^2(Y_{t-1}, Y_t). \tag{14}$$

Now, since the square of the overall correlation is upper bounded by $P_{ij}^{(r)}$, then our previous remarks on ergodic chains result in

Theorem 3.1 For a bijective round function $F : Z_2^n \to Z_2^n$ with an ergodic correlation matrix $\mathbf{P}$, the correlation between any linear function of the plaintext Y_0 and any linear function of the ciphertext Y_r for a large number of rounds r, asymptotically satisfies

$$|2q^* - 1| = \left|\prod_{i=1}^{r} c(Y_{i-1}, Y_i)\right| \leq \frac{1}{\sqrt{N}}. \tag{15}$$

□

As before we pass to random graph theory to prove the irreducibility of $\mathbf{P}$, and hence its ergodicity. Because of the combinatorial structure in the correlation matrix it is possible to prove that almost all entries in $\mathbf{P}$ are nonzero, and that the underlying graph $\mathbf{G}$ is tending to the complete graph.

Theorem 3.2 (O'Connor and Golić [11]) For uniformly distributed bijective function F the probability that the correlation matrix is irreducible is $1 - o(N^{-1})$ when $N = 2^n - 1$. □

4 Conclusion

The Markov approach to product cipher design has been discussed with respect to two cryptographic attacks. The convergence of these chains to the uniform distribution suggests that the iterative structure of product ciphers is a good design principle for generating mappings for which certain measures of strength should tend to be distributed uniformly. There are several difficulties with the Markov approach, one of which the reader should perceive is related to the ergodicity of the chains defined. To demonstrate that $\mathbf{P}$ is ergodic we must show that it is finite, aperiodic and irreducible, and of these conditions only the last is nontrivial to demonstrate. Generally speaking, the process will be irreducible if it can be shown that $\mathbf{P}$ has a 'large' number of nonzero entries. In particular, if $\mathbf{P}$ is $N \times N$ then having slightly more than $N \log N$ edges will suffice with high probability. This bound is taken from random graph theory, and is the threshold function for a directed graph to be strongly connected.

However, in most cases the dimensions of $\mathbf{P}$ will make a direct inspection of its entries infeasible. For DES the $\mathbf{P}$ matrix has about $2^{64} \times 2^{64}$ entries. This then suggests that some probabilistic approach must be taken to proving the irreducibility of $\mathbf{P}$. We can prove results for a random round function F or even 'almost all' round functions but it appears difficult to say something meaningful about a given round function, say that of DES. A new design principle for round functions is to guarantee the irreducibility property in the $\mathbf{P}$ matrix. If this criterion was to be adopted, it would be useful to determine the consequences for other statistical tests of security in a product cipher if the differential and linear cryptanalysis chains tend to be uniform. For example, the distribution of differences effects the strict avalanche criterion, which refers to changes in the ciphertext in response to small changes in the plaintext. It may be the case that guaranteeing a close to uniform spread in the differences causes most other statistical tests to be close to optimal as well.

Probably the main limitation to the Markov approach is the difficulty in readily determining the convergence of the chain to its stationary distribution. Standard arguments

(see [2] for example) state that if $P_{ij}^{(r)} = \pi_i + e_{ij}^{(r)}$ then

$$|e_{ij}^{(r)}| \leq cr^n \tag{16}$$

which states that the convergence is geometric. This bound on the error terms is conservative and is derived from the size of the smallest element in $\mathbf{P}$, which for the chains discussed here, does not yield a useful bound. More advanced techniques, possibly based on eigenvalues, are required to determine the convergence more accurately. A lower bound on the rate of convergence has been determined by O'Connor [9] for differentially 2-uniform mappings.

References

[1] R. B. Ash. *Information Theory.* New York: Dover Publications, 1965.

[2] U. Bhat. *Elements in applied stochastic processes.* John Wiley and Sons, 1972.

[3] E. Biham and A. Shamir. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, 4(1):3–72, 1991.

[4] B. Bollobás. *Random graphs.* Academic Press, 1985.

[5] R. G. Gallager. *Low Density Parity Check Codes.* MIT Press, Cambridge. Mass., 1963.

[6] M. Matsui. Linear cryptanalysis method for DES cipher. *abstracts of papers, EUROCRYPT 93, Norway, May.*

[7] M. Matsui. Linear cryptanalysis of DES cipher (I). private communication.

[8] H. Minc. *Nonnegative matrices.* John Wiley and Sons, 1988.

[9] L. J. O'Connor. Convergence in differential distributions. submitted.

[10] L. J. O'Connor. On the distribution of characteristics in bijective mappings. presented at Eurocrypt 93, Norway, May 1993. Also accepted for publication in the *Journal of Cryptology.*

[11] L. J. O'Connor and J. Dj Golić. A Markov approach to Linear Cryptanalysis. submitted.

[12] I. Palásti. On the strong connectedness of random graphs. *Studia Sci. Math. Hungar.*, 1:205–214, 1966.

[13] R. A. Rueppel. Stream ciphers. In G. Simmons, editor, *Contemporary Cryptology: the Science of Information Integrity*, pages 64–134. IEEE Press, 1991.

[14] U. Vazirani. Rapidly mixing Markov chains. In B. Bollobás, editor, *Probabilistic combinatorics and its applications, proceedings of Symposia in Applied Mathematics, volume 44*, pages 99–121, 1991.

Nonlinearity Characteristics of Quadratic Substitution Boxes *

Jennifer Seberry
Xian-Mo Zhang
Yuliang Zheng

Department of Computer Science
The University of Wollongong
Wollongong, NSW 2522, AUSTRALIA
E-mail: {jennie,xianmo,yuliang}@cs.uow.edu.au

January 25, 1994

Abstract

An important question in designing cryptographic functions including substitution boxes (S-boxes) is the relationships among the various nonlinearity criteria each of which indicates the strength or weakness of a cryptographic function against a particular type of cryptanalytic attacks. In this paper we reveal, for the first time, interesting connections among the strict avalanche characteristics, differential characteristics, linear structures and nonlinearity of quadratic S-boxes. In addition, we show that our proof techniques allow us to treat in a unified fashion all quadratic permutations, regardless of the underlying construction methods. This greatly simplifies the proofs for a number of known results on nonlinearity characteristics of quadratic permutations. As a by-product, we obtain a negative answer to an open problem regarding the existence of differentially 2-uniform quadratic permutations on an even dimensional vector space.

1 Nonlinearity Criteria

We first introduce basic notions and definitions of several nonlinearity criteria for cryptographic functions.

Denote by V_n the vector space of n tuples of elements from $GF(2)$. Let $\alpha = (a_1, \ldots, a_n)$ and $\beta = (b_1, \ldots, b_n)$ be two vectors in V_n. The scalar product of α and β, denoted by $\langle \alpha, \beta \rangle$, is defined by $\langle \alpha, \beta \rangle = a_1 b_1 \oplus \cdots \oplus a_n b_n$, where multiplication and addition are over $GF(2)$. In this paper we consider functions from V_n to $GF(2)$ (or simply functions on V_n). We are particularly interested in functions whose algebraic degrees are 2, also called quadratic functions. These functions take the form of $a_{00} \oplus \sum_{1 \leq i,j \leq n} a_{ij} x_i x_j$, where a_{ij} is an element from $GF(2)$, while x_i is a variable in $GF(2)$.

Let f be a function on V_n. The $(1,-1)$-sequence defined by $((-1)^{f(\alpha_0)}, (-1)^{f(\alpha_1)}, \ldots, (-1)^{f(\alpha_{2^n-1})})$ is called the *sequence* of f, and the $(0,1)$-sequence defined by $(f(\alpha_0), f(\alpha_1),$

*The first author was supported in part by the Australian Research Council under the reference numbers A49130102, A9030136, A49131885 and A49232172, the second author by A49130102, and the third author by A49232172.

$\ldots, f(\alpha_{2^n-1}))$ is called the *truth table* of f, where $\alpha_0 = (0,\ldots,0,0)$, $\alpha_1 = (0,\ldots,0,1)$, $\ldots$, $\alpha_{2^n-1} = (1,\ldots,1,1)$. f is said to be *balanced* if its truth table has 2^{n-1} zeros (ones).

An *affine* function f on V_n is a function that takes the form of $f = a_1x_1 \oplus \cdots \oplus a_nx_n \oplus c$, where $a_j, c \in GF(2), j = 1,2,\ldots,n$. Furthermore f is called a *linear* function if $c = 0$. The sequence of an affine (or linear) function is called an *affine (or linear) sequence.*

The *Hamming weight* of a vector $\alpha \in V_n$, denoted by $W(\alpha)$, is the number of ones in the vector.

Now we introduce bent functions, an important combinatorial concept introduced by Rothaus in the mid 1960's (although his pioneering work was not published until some ten years later [17].)

Definition 1 *A function f on V_n is said to be bent if*

$$2^{-\frac{n}{2}} \sum_{x \in V_n} (-1)^{f(x) \oplus \langle \beta, x \rangle} = \pm 1$$

for every $\beta \in V_n$. Here $x = (x_1,\ldots,x_n)$ and $f(x) \oplus \langle \beta, x \rangle$ is considered as a real valued function.

¿From the definition, it can be seen that bent functions on V_n exist only when n is even. Another fact is that bent functions are not balanced, hence not directly applicable in most computer and communications security practices. Dillon presented a nice exposition of bent functions in [7]. In particular, he showed that bent functions can be characterized in various ways:

Lemma 1 *The following statements are equivalent:*

(i) f is bent.

(ii) $\langle \xi, \ell \rangle = \pm 2^{\frac{1}{2}n}$ for any affine sequence ℓ of length 2^n, where ξ is the sequence of f.

(iii) $f(x) \oplus f(x \oplus \alpha)$ is balanced for any non-zero vector $\alpha \in V_n$, where $x = (x_1,\ldots,x_n)$.

The strict avalanche criterion (SAC) was first introduced by Webster and Tavares [23, 24] when studying the design of cryptographically strong substitution boxes (S-boxes).

Definition 2 *A function f on V_n is said to satisfy the strict avalanche criterion (SAC) if $f(x) \oplus f(x \oplus \alpha)$ is balanced for all $\alpha \in V_n$ with $W(\alpha) = 1$, where $x = (x_1,\ldots,x_n)$.*

It is widely accepted that the component functions of an S-box employed by a modern block cipher should all satisfy the SAC. A general technique for constructing SAC-fulfilling cryptographic functions can be found in [21].

While the SAC measures the avalanche characteristics of a function, the linear structure is a concept that in a sense complements the former, namely, it indicates the straightness of a function.

Definition 3 *Let f be a function on V_n. A vector $\alpha \in V_n$ is called a* linear structure *of f if $f(x) \oplus f(x \oplus \alpha)$ is a constant.*

Evertse apparently was the first person who studied implications of linear structures (in a sense broader than ours) on the security of encryption algorithms [8]. By definition, the zero vector in V_n is a linear structure of all functions on V_n. It is not hard to see that the linear structures of a function f form a linear subspace of V_n. The dimension of the subspace is called the *linearity dimension* of f. Clearly, the linearity dimension of a function on V_n is bounded from the above by n, with the affine functions achieving the maximum dimension n. It is bounded from the below by 0 when n is even and by 1 when n is odd. The lower bound 0 is achieved only by bent functions that have the zero vector as their only linear structure, while 1 can be achieved by functions that have only two linear structures (one is the zero vector and the other is a nonzero vector). Examples of the latter are those obtained by concatenating two bent functions (see [18, 22]).

In mathematical terms, an $n \times s$ S-box (i.e., with n input bits and s output bits), can be described as a mapping from V_n to V_s ($n \geqq s$). To avoid trivial statistical attacks, an S-box F should be *regular*, namely, $F(x)$ should run through all vectors in V_s each 2^{n-s} times while x runs through V_n once. Note that an $n \times n$ S-box is a permutation on V_n and always regular.

Regularity of an $n \times s$ S-box F can be characterized by the balance of nonzero linear combinations of its component functions. It has been known that when $n = s$, F is regular if and only if all nonzero linear combinations of the component functions are balanced. A proof can be found in Remark 5.8 of [7]. The characterization can be extended to the case when $n > s$.

Theorem 1 *Let $F = (f_1, \ldots, f_s)$, where f_i is a function on V_n, $n \geqq s$. Then F is a regular mapping from V_n to V_s if and only if all nonzero linear combinations of $f_1, \ldots, f_n$ are balanced.*

A proof for the theorem is given in Appendix A. It seems to the authors that the proof for the case of $n = s$ as described in [7] can not be directly adapted to the general case of $n > s$, and hence the extension presented here is not trivial.

The next criterion is the nonlinearity that indicates the Hamming distance between a function and all the affine functions.

Definition 4 *Given two functions f and g on V_n, the* Hamming distance *between them, denoted by $d(f, g)$, is defined as the Hamming weight of the truth table of the function $f(x) \oplus g(x)$, where $x = (x_1, \ldots, x_n)$. The* nonlinearity *of f, denoted by N_f, is the minimal Hamming distance between f and all affine functions on V_n, i.e., $N_f = \min_{i=1,2,\ldots,2^{n+1}} d(f, \varphi_i)$ where φ_1, φ_2, ..., $\varphi_{2^{n+1}}$ denote the affine functions on V_n.*

The above definition can be extended to the case of mappings, by defining the nonlinearity of a mapping from V_n to V_s as the minimum among the nonlinearities of nonzero linear combinations of the component functions.

The nonlinearity of a function f on V_n has been known to be bounded from the above by $2^{n-1} - 2^{\frac{1}{2}n-1}$. When n is even, the upper bound is achieved by bent functions. Constructions for highly nonlinear *balanced* functions can be found in [18, 22].

Nonlinearity has been considered to be an important criterion. Recent advances in *Linear cryptanalysis* put forward by Matsui [10] have made it explicit that nonlinearity is not just important, but essential to DES-like block encryption algorithms. Linear cryptanalysis exploits the low nonlinearity of S-boxes employed by a block cipher, and it has been successfully applied in attacking FEAL and DES. In [20], it has been shown that to immunize

an S-box against linear cryptanalysis, it suffices for the Hamming distance between each nonzero linear combination of the component functions and each affine function not to deviate too far from 2^{n-1}, namely, *an S-box is immune to linear cryptanalysis if the nonlinearity of each nonzero linear combination of its component functions is high.*

Finally we consider a nonlinearity criterion that measures the strength of an S-box against differential cryptanalysis [3, 4]. The essence of a differential attack is that it exploits particular entries in the difference distribution tables of S-boxes employed by a block cipher. The difference distribution table of an $n \times s$ S-box is a $2^n \times 2^s$ matrix. The rows of the matrix, indexed by the vectors in V_n, represent the change in the input, while the columns, indexed by the vectors in V_s, represent the change in the output of the S-box. An entry in the table indexed by (α, β) indicates the number of input vectors which, when changed by α (in the sense of bit-wise XOR), result in a change in the output by β (also in the sense of bit-wise XOR).

Note that an entry in a difference distribution table can only take an even value, the sum of the values in a row is always 2^n, and the first row is always $(2^n, 0, \ldots, 0)$. As entries with higher values in the table are particularly useful to differential cryptanalysis, a necessary condition for an S-box to be immune to differential cryptanalysis is that it does not have large values in its differential distribution table (not counting the first entry in the first row).

Definition 5 *Let F be an $n \times s$ S-box, where $n \geqq s$. Let δ be the largest value in differential distribution table of the S-box (not counting the first entry in the first row), namely,*

$$\delta = \max_{\alpha \in V_n, \alpha \neq 0} \max_{\beta \in V_s} |\{x | F(x) \oplus F(x \oplus \alpha) = \beta\}|.$$

Then F is said to be differentially δ-uniform, *and accordingly, δ is called the differential uniformity of f.*

Obviously the differential uniformity δ of an $n \times s$ S-box is constrained by $2^{n-s} \leqq \delta \leqq 2^n$. Extensive research has been carried out in constructing differentially δ-uniform S-boxes with a low δ [12, 1, 13, 15, 14, 2]. Some constructions, in particular those based on permutation polynomials on finite fields, are simple and elegant. However, cautions must be taken with Definition 5. In particular, it should be noted that low differential uniformity (a small δ) is only a *necessary*, but not a *sufficient* condition for immunity to differential attacks. This is shown by the fact that S-boxes constructed in [12, 1] are extremely weak to differential attacks, despite that they achieve the lowest possible differential uniformity $\delta = 2^{n-s}$ [4, 5, 20]. A more complete measurement is the *robustness* introduced in [20]. The reader is directed to that paper for a comprehensive treatment of this subject.

Note that an $n \times s$ S-box achieves the lowest possible differential uniformity $\delta = 2^{n-s}$ if and only if it has a *flat* difference distribution table. As has been noticed by many researchers (see for instance Page 62 of [4]), a flat difference distribution table is not associated with a regular S-box. This result, together a formal proof, is reviewed in the following.

Lemma 2 *The differential uniformity of a regular $n \times s$ S-box is larger than 2^{n-s}.*

Proof. Let F is a regular $n \times s$ S-box. By Theorem 1, nonzero linear combinations of the component functions of F are all balanced. Assume for contradiction that for each nonzero $\alpha \in V_n$, $F(x) \oplus F(x \oplus \alpha)$ is regular, namely it runs through all vectors in V_s, each 2^{n-s} times,

while x runs through V_n once. Recall that Theorem 3.1 of [12] states that $F(x) \oplus F(x \oplus \alpha)$ is regular if and only if each nonzero linear combination of the component functions of F is a bent function. This contradicts the fact that each nonzero linear combination of the component functions of F is balanced. □

We have discussed various cryptographic properties including the algebraic degree, the SAC, the linear structure, the regularity, the nonlinearity and the differential uniformity. As is stated in the following lemmas, some properties are invariant under a nonsingular linear transformation.

Lemma 3 *Let f be a function on V_n, A be a nonsingular matrix of order n over $GF(2)$, and let $g(x) = f(xA)$. Then f and g have the same algebraic degree, nonlinearity and linearity dimension.*

Proof. The algebraic degree of a function is obviously not changed by a nonsingular affine transformation on input coordinates. The invariance of nonlinearity was pointed out in [11], while that of linearity dimension follows from the fact that linear structures form a subspace whose dimension remains the same under the transformation. □

The next lemma was pointed out in Section 5.3 of [20]. It was also noticed by Beth and Ding in [2]. The lemma is followed by a short formal proof for the sake of completeness.

Lemma 4 *Let F be a mapping from V_n to V_s, where $n \geqq s$, A be a nonsingular matrix of order n over $GF(2)$, and B be a nonsingular matrix of order s over $GF(2)$. Let $G(x) = F(xA)$ and $H(x) = F(x)B$, where $x = (x_1, \ldots, x_n)$. Note that A is applied to the input, while B to the output of F. Then F, G and H all have the same regularity and differential uniformity.*

Proof. Let β be a vector in V_s. Since $F(x) = G(xA^{-1})$, $F(x) = \beta$ if and only if $G(xA^{-1}) = \beta$. This implies that, while x runs through V_n, $F(x)$ and $G(x)$ run through β the same number of times.

Now consider $H(x) = F(x)B$. Clearly $F(x) = \beta$ if and only if $H(x) = F(x)B = \beta B$. As B is nonsingular, $F(x)$ runs through β exactly the same number of times as that $H(x)$ runs through βB, while x runs through V_n. □

2 Cryptographic Properties of Quadratic S-boxes

In this section we reveal interesting relationships among the difference distribution table, linear structures, nonlinearity and SAC of S-boxes whose component functions are all quadratic (or simply, quadratic S-boxes).

2.1 Linear Structure vs Nonlinearity

Consider a quadratic function f on V_n. Then $f(x) \oplus f(x \oplus \alpha)$ is affine, where $x = (x_1, \ldots, x_n)$ and $\alpha \in V_n$. Assume that f does not have nonzero linear structures. Then for any nonzero $\alpha \in V_n$, $f(x) \oplus f(x \oplus \alpha)$ is a nonzero affine function, hence balanced. By Part (iii) of Lemma 1, f is bent. Thus we have:

Lemma 5 *If a quadratic function f on V_n has no nonzero linear structures, then f is bent and n is even.*

The following lemma is a useful tool in calculating the nonlinearity of functions obtained via Kronecker product.

Lemma 6 *Let $g(x,y) = f_1(x) \oplus f_2(y)$, where $x = (x_1, \ldots, x_{n_1})$, $y = (y_1, \ldots, y_{n_2})$, f_1 is a function on V_{n_1} and f_2 is a function on V_{n_1}. Let d_1 and d_2 denote the nonlinearities of f_1 and f_2 respectively. Then the nonlinearity of g satisfies*

$$N_g \geqq d_1 2^{n_2} + d_2 2^{n_1} - 2d_1 d_2.$$

In addition, we have $N_g \geqq d_1 2^{n_2}$ and $N_g \geqq d_2 2^{n_1}$.

Proof. The first half of the lemma can be found in Lemma 8 of [19]. The second half is true due to the fact that $d_1 \leqq 2^{n_1-1}$ and $d_2 \leqq 2^{n_2-1}$ (see also Section 3 of [18]). □

We now examine how the nonlinearity of a function on V_n relates to the linearity dimension of the function.

Let g be a (not necessarily quadratic) function on V_n, $\{\beta_1, \ldots, \beta_\ell\}$ be a basis of the subspace consisting of the linear structures of g. $\{\beta_1, \ldots, \beta_\ell\}$ can be extended to $\{\beta_1, \ldots, \beta_\ell, \beta_{\ell+1}, \ldots, \beta_n\}$ such that the latter is a basis of V_n. Now let B be a nonsingular matrix with β_i as its ith row, and let $g^*(x) = g(xB)$. By Lemma 3, g^* and g have the same linearity dimension, algebraic degree and nonlinearity. Thus the question is transformed into the discussion of g^*.

Let e_i be the vector in V_n whose ith coordinate is one and others are zero. Then we have $e_j B = \beta_j$, and $g^*(e_i) = g(\beta_i)$, $i = 1, \ldots, n$. Thus $\{e_1, \ldots, e_\ell\}$ is a basis of the subspace consisting of the linear structures of g^*. Write g^* as

$$g^*(x) = q(y) \oplus \sum_j [m_j(y) r_j(z)] \tag{1}$$

where $x = (x_1, \ldots, x_n)$, $y = (x_1, \ldots, x_\ell)$, $z = (x_{\ell+1}, \ldots, x_n)$, $m_j \neq 0$, the algebraic degree of each r_j is at least 1 and $r_j \neq r_i$ for $j \neq i$. Also write e_i as $e_i = (\mu_i, 0)$, where $\mu_j \in V_\ell$ and $0 \in V_{n-\ell}$. As e_i is a linear structure of g^*, the following difference

$$g^*(x) \oplus g^*(x \oplus e_i) = q(y) \oplus q(y \oplus \mu_i) \oplus \sum_j [(m_j(y) \oplus m_j(y \oplus \mu_i)) r_j(z)]$$

is a constant. This implies that $q(y) \oplus q(y \oplus \mu_i)$ is a constant (i.e. μ_i is a linear structure of $q(y)$) and each $m_j(y) \oplus m_j(y \oplus \mu_j) = 0$ (i.e. $m_j = 1$). Thus (1) can be rewritten as

$$g^*(x) = q(y) \oplus r(z). \tag{2}$$

Since all vectors in V_ℓ are linear structures of q, q is an affine function on V_ℓ. As the linearity dimension of g^* is also ℓ, r must be a function on $V_{n-\ell}$ that does *not* have nonzero linear structures. By Lemmas 3 and 6, we have $N_g = N_{g^*} \geqq 2^\ell N_r$. This is precisely what Proposition 3 of [13] states.

As a special case, suppose that g in the above discussions is quadratic. Then the function r in (2) is a quadratic function on $V_{n-\ell}$ with no nonzero linear structures. By Lemma 5, r is a bent function on $V_{n-\ell}$ whose nonlinearity is $N_r = 2^{n-\ell-1} - 2^{\frac{1}{2}(n-\ell)-1}$. Thus we have:

Theorem 2 *Let g be a function on V_n whose algebraic degree is at most 2. Denote by ℓ the linearity dimension of g. Then*

(i) $n-\ell$ is even, and

(ii) the nonlinearity of g satisfies $N_g \geqq 2^{n-1} - 2^{\frac{1}{2}(n+\ell)-1}$.

The lower bound on nonlinearity in Theorem 2 can be straightforwardly translated into that for quadratic (not necessarily regular) $n \times s$ S-boxes ($n \geqq s$).

Now we take a closer look at the nonlinearity of a quadratic function g on V_n whose linearity dimension is ℓ. As g is nonlinear, we have $\ell < n$. In addition since g is quadratic, by (i) of Theorem 2, $n-\ell$ is even. Thus we have $\ell \leqq n-2$, and $N_g \geqq 2^{n-1} - 2^{\frac{1}{2}(n+\ell)-1} \geqq 2^{n-2}$. This proves the following:

Corollary 1 *The nonlinearity of a quadratic function on V_n is at least 2^{n-2}.*

Corollary 1 is a bit surprising in the sense that it indicates that all quadratic functions are fairly nonlinear, and there is no quadratic function whose nonlinearity is between 0 and 2^{n-2} (exclusive).

2.2 Difference Distribution Table vs Linear Structure

First we show an interesting result stating that the number representing the differential uniformity of a quadratic S-box must be a power of 2.

Theorem 3 *Let δ be the differential uniformity of a quadratic $n \times s$ S-box. Then $\delta = 2^d$ for some $n-s \leqq d \leqq n$. Furthermore, if the S-box is regular, then we have $\delta = 2^d$ for some $n-s+1 \leqq d \leqq n$.*

Proof. Let $F = (f_1, \ldots, f_s)$. Let α be a nonzero vector in V_n. Then

$$F(x) \oplus F(x \oplus \alpha) = (f_1(x) \oplus f_1(x \oplus \alpha), \ldots, f_s(x) \oplus f_s(x \oplus \alpha)).$$

As f_i is quadratic, $f_i(x) \oplus f_i(x \oplus \alpha)$ is affine, hence $F(x) \oplus F(x \oplus \alpha) = xD \oplus c$, where D is an $n \times s$ matrix over $GF(2)$ and C is a vector in V_s.

Assume that the rank of D is r with $0 \leqq r \leqq s$. Then $F(x) \oplus F(x \oplus \alpha) = xD \oplus C$ runs through 2^r vectors in V_s, each 2^{n-r} times, while x runs through V_n, where n, s and r satisfy $n-s \leqq n-r \leqq n$. Thus the differential uniformity of F takes the form of 2^d, $n-s \leqq d \leqq n$.

The second half of the lemma follows from Lemma 2 together with the above discussions. □

Let $F = (f_1, \ldots, f_s)$ be a regular quadratic $n \times s$ S-box, and let g be a nonlinear combination of the component functions of F. Then it can be shown that g has at least one nonzero linear structure. To prove the claim, we assume that g has no nonzero linear structures. Then by Lemma 5, g is a bent function. This contradicts the fact that F is regular and that the nonzero linear combinations of its component functions are all balanced and have linear structures.

Next we show that the differential uniformity of an S-box is closely related to the number of linear structures of an nonzero linear combinations of the component functions of the S-box.

Theorem 4 *Let $F = (f_1, \ldots, f_s)$ be a regular quadratic $n \times s$ S-box. Then the differential uniformity of F satisfies $\delta \leqq 2^{n-s+t}$, where $1 \leqq t \leqq s$ (see also Theorem 3), if and only if any nonzero vector $\alpha \in V_n$ is a linear structure of at most $2^t - 1$ nonzero linear combinations of $f_1, \ldots, f_s$.*

Proof. (i) First we show that if $\delta \leqq 2^{n-s+t}$, then any nonzero vector $\alpha \in V_n$ is a linear structure of at most $2^t - 1$ nonzero linear combinations of the component functions. To simplify our proof, it can be assumed that $\delta = 2^{n-s+t}$.

Note that there are $2^s - 1$ nonzero linear combinations of $f_1, \ldots, f_s$, denoted by $g_1, \ldots, g_{2^s-1}$, and $2^n - 1$ nonzero vectors in V_n, denoted by $\alpha_1, \ldots, \alpha_{2^n-1}$. Now suppose that there exist 2^t nonzero linear combinations $g_1, \ldots, g_{2^t}$, such that α is a linear structure of each g_j. Write $g_j(x) \oplus g_j(x \oplus \alpha) = a_j$, where a_j is constant, $j = 1, \ldots, 2^t$. Let $\Omega = \{g_1, \ldots, g_{2^t}\}$. We are interested in the rank of Ω, namely the maximum number of functions in Ω that are linearly independent. Recall that t linearly independent functions can generate only $2^t - 1$ distinct nonzero combinations. As Ω contains 2^t nonzero functions, its rank is at least $t+1$. Without loss of generality, suppose that $g_1, \ldots, g_{t+1}$ are linearly independent. Then there exist additional $s - t - 1$ nonzero linear combinations of $f_1, \ldots, f_s$, denoted by $h_{t+2}, \ldots, h_s$, such that $g_1, \ldots, g_{t+1}, h_{t+2}, \ldots, h_s$ are all linearly independent. Let G be an $n \times s$ mapping defined by $G = (g_1, \ldots, g_{t+1}, h_{t+2}, \ldots, h_s)$. Then G can be expressed as $G(x) = F(x)B$ for a nonsingular matrix B of order s over $GF(2)$.

By Lemma 4, G is also a differentially δ-uniform $n \times s$ S-box. Since $\delta = 2^{n-s+t}$ $(1 \leqq t \leqq s)$, $G(x) \oplus G(x \oplus \alpha)$ runs through at least $2^n/2^{n-s+t} = 2^{s-t}$ vectors. On the other hand,

$$G(x) \oplus G(x \oplus \alpha) = (a_1, \ldots, a_{t+1}, h_{t+2}(x) \oplus h_{t+2}(x \oplus \alpha), \ldots, h_s(x) \oplus h_s(x \oplus \alpha))$$

where $a_1, \ldots, a_{t+1}$ are all constants. This indicates that $G(x) \oplus G(x \oplus \alpha)$ runs through at most 2^{s-t-1} vectors in V_s. This is a contradiction.

(ii) Next we prove the other direction. Suppose any nonzero vector $\alpha \in V_n$ is a linear structure of m nonzero linear combinations of the component functions, where $m \leqq 2^t - 1$. We show that the differential uniformity of F is at most 2^{n-s+t}.

Let $W = \{g_1, \ldots, g_{2^s-1}\}$ be the set of the $2^s - 1$ nonzero linear combinations of $f_1, \ldots, f_s$. Let α be a nonzero vector in V_n. Note that together with the zero function, the functions in W which have α as their linear structure form a linear space. The number of nonzero functions in the set is $2^{t'} - 1$ for some $t' < t$. Without loss of generality, let $U_\alpha = \{g_1, \ldots, g_{2^{t'}-1}\}$ be the set of nonzero functions having α as their linear structure.

Since U_α, together with the zero function, forms a linear space, it contains precisely t' linearly independent functions. Without loss of generality, let $g_1, \ldots, g_{t'}$ be t' linearly independent functions. Now let $h_{t'+1}, \ldots, h_s$ be $s - t'$ additional nonzero linear combinations of $f_1, \ldots, f_s$, such that $g_1, \ldots, g_{t'}, h_{t'+1}, \ldots, h_s$ are all linearly independent. Set $G = (g_1, \ldots, g_{t'}, h_{t'+1}, \ldots, h_s)$. As α is a linear structure of g_j, $j = 1, \ldots, t'$, we have

$$G(x) \oplus G(x \oplus \alpha) = (a_1, \ldots, a_{t'}, h_{t'+1}(x) \oplus h_{t'+1}(x \oplus \alpha) \ldots h_s(x) \oplus h_s(x \oplus \alpha)) \qquad (3)$$

where each a_j, $j = 1, \ldots, t'$, is a constant.

We now show that $h_{t'+1}(x) \oplus h_{t'+1}(x \oplus \alpha), \ldots, h_s(x) \oplus h_s(x \oplus \alpha)$ are linearly independent. Suppose that $h_{t'+1}(x) \oplus h_{t'+1}(x \oplus \alpha), \ldots, h_s(x) \oplus h_s(x \oplus \alpha)$ are linearly dependent. Then there exists a nonzero vector $(c_{t'+1}, \ldots, c_s) \neq (0, \ldots, 0)$ such that

$$\sum_{j=t'+1}^{s} c_j[h_j(x) \oplus h_j(x \oplus \alpha)] = 0. \qquad (4)$$

Write $h(x) = \sum_{j=t'+1}^{s} c_j h_j(x)$. h is a nonzero function as $h_{t'+1}, \ldots, h_s(x)$ are linearly independent. Thus (4) implies that α is a nonzero linear structure of h. In other words, we have $h \in U_\alpha$. On the other hand, since $g_1, \ldots, g_{t'}, g_{t'+1}, \ldots, g_s$ are linearly independent, we have $h \notin U_\alpha$. The above contradiction shows that $h_{t'+1}(x) \oplus h_{t'+1}(x \oplus \alpha), \ldots, h_s(x) \oplus h_s(x \oplus \alpha)$ are indeed linearly independent.

As the $s - t'$ affine functions $h_{t'+1}(x) \oplus h_{t'+1}(x \oplus \alpha), \ldots, h_s(x) \oplus h_s(x \oplus \alpha)$ are linearly independent, $G(x) \oplus G(x \oplus \alpha)$ runs through $2^{s-t'}$ vectors in V_n, each $2^n/2^{s-t'} = 2^{n-s+t'}$ times. Hence the differential uniformity of $G(x)$ satisfies $\delta = 2^{n-s+t'} \leqq 2^{n-s+t}$. By Lemma 4, $F(x)$ and $G(x)$ have an identical differential uniformity. □

Theorem 4 indicates that with an S-box with a smaller δ, i.e., a smaller t, the nonzero linear combinations of its component functions have less linear structures. This coincides with our intuition that the nonlinearity of an S-box grows with the strength of its immunity to differential attacks.

2.3 Difference Distribution Table vs SAC

Armed with Theorem 4, we further reveal that differential uniformity is tightly associated with the strict avalanche characteristics.

Theorem 5 *Let $F = (f_1, \ldots, f_s)$ be a differentially δ-uniform regular quadratic $n \times s$ S-box, where $\delta = 2^{n-s+t}$, $1 \leqq t \leqq s$ (see also Theorem 3). If t and s satisfy $s \leqq 2^{s-t-2}$, then there exists a nonsingular matrix of order n over $GF(2)$, say A, and a nonsingular matrix of order s over $GF(2)$, say B, such that $\Psi(x) = F(xA)B = (f_1(xA), \ldots, f_s(xA))B = (\psi_1(x), \ldots, \psi_s(x))$ is also a differentially δ-uniform regular quadratic $n \times s$ S-box whose component functions all satisfy the SAC.*

Proof. Again denote by $g_1, \ldots, g_{2^s-1}$ the $2^s - 1$ nonzero linear combinations of $f_1, \ldots, f_s$, and by $\alpha_1, \ldots, \alpha_{2^n-1}$ the $2^n - 1$ nonzero vectors in V_s. We construct a bipartite graph Γ with $g_1, \ldots, g_{2^s-1}$ on one side and $\alpha_1, \ldots, \alpha_{2^n-1}$ on the other side. An edge exists between g_i and α_j if and only if α_j is a linear structure of g_i. By Theorem 4, there exist at most $2^t - 1$ edges associated with each α. Thus there exist at most $(2^t - 1) \cdot (2^n - 1)$ edges in the graph Γ.

Denote by t_j the number of linear structures of g_j, $j = 1, \ldots, 2^s - 1$. Without loss of generality suppose that $t_1 \leqq t_2 \leqq \cdots \leqq t_{2^s-1}$. It can be seen that $t_j < 2^{n-s+t+1}$, $j = 1, \ldots, 2^{s-1}$. The reason is as follows. Suppose that it is not the case. Then we have $t_1 + \cdots + t_{2^s-1} \geqq 2^{s-1} \cdot 2^{n-s+t+1} = 2^{n+t} > (2^t - 1) \cdot (2^n - 1)$. This contradicts the fact that Γ has at most $2^{t-1} \cdot (2^n - 1)$ edges.

Now set $\Omega = \{g_1, \ldots, g_{2^{s-1}+1}\}$. As the rank of Ω is s, we can choose s functions from Ω, say $g_{j_1}, \ldots, g_{j_s}$, such that they are all linearly independent. Since $s \leqq 2^{s-t-2}$, we have $t_{j_1} + \cdots + t_{j_s} < s \cdot 2^{n-s+t+1} \leqq 2^{n-1}$. By Theorem 2 of [21], there exists a nonsingular matrix A of order n over $GF(2)$, such that all component functions of $(g_{j_1}(xA), \ldots, g_{j_s}(xA))$ satisfy the SAC. Furthermore, as each g_j is a nonzero linear combination of $f_1, \ldots, f_s$, there is a nonsingular matrix B of order s over $GF(2)$ such that $(g_{j_1}(x), \ldots, g_{j_s}(x)) = (f_1(x), \ldots, f_s(x))B$. Accordingly, by Lemma 4,

$$\Psi(x) = F(xA)B = (f_1(xA), \ldots, f_s(xA))B = (\psi_1(x), \ldots, \psi_s(x))$$

is a differentially δ-uniform regular quadratic $n \times s$ S-box, where each component function ψ_j satisfies the SAC. □

In Theorem 5, when the differential uniformity $\delta = 2^{n-s+t}$ is small, the parameter t is also small, and the condition $s \leqq 2^{s-t-2}$ is likely to be satisfied. Equivalently we can say that S-boxes strong against differential attacks are also SAC-fulfilling, subject to a nonsingular linear transformation. Again, this coincides with our intuition.

3 A Unified Treatment of Quadratic Permutations

This section is concerned with differentially 2-uniform quadratic $n \times n$ S-boxes. Since such an S-box F is a permutation, $F(x) \oplus F(x \oplus \alpha)$ takes a vector two times or does not take it, while x runs through V_n once. F has the following property: for any nonzero vector $\alpha \in V_n$, $F(x) \oplus F(x \oplus \alpha)$ runs through 2^{n-1} vectors in V_n, each twice, but not through the other 2^{n-1} vectors, while x runs through V_n.

Although there are many question marks regarding the applicability of differentially 2-uniform quadratic $n \times n$ S-boxes in computer security practices, primarily due to their low algebraic degree, these S-boxes have received extensive research in the past years [16, 15, 6, 2, 14] and hence deserve our special attention. These S-boxes appear in various forms and researchers have employed different techniques, some of which are rather sophisticated, to prove their nonlinearity. By refining our proof techniques described in Section 2, we will show in this section that all differentially 2-uniform quadratic permutations, no matter how they are constructed, have the same nonlinearity and can be transformed into SAC-fulfilling S-boxes. This greatly simplifies the proof for a number of known results and could be a powerful tool in designing cryptographically strong block ciphers.

Theorem 6 *Let $F = (f_1, \ldots, f_n)$ be a quadratic permutation on V_n. Then the following statements are equivalent:*

(i) for any nonzero linear combination of $f_1, \ldots, f_n$, say $g = \sum_{j=1}^{n} c_j f_j$, its nonlinearity satisfies $N_g \geqq 2^{n-1} - 2^{\frac{1}{2}(n-1)}$.

(ii) any nonzero linear combination of $f_1, \ldots, f_n$, say $g = \sum_{j=1}^{n} c_j f_j$, has a unique nonzero linear structure.

(iii) each nonzero vector in V_n is the linear structure of a unique nonzero linear combination of $f_1, \ldots, f_n$.

(iv) F is differentially 2-uniform, i.e. for each nonzero vector $\alpha \in V_n$, $F(x) \oplus F(x \oplus \alpha)$ runs through half of the vectors in V_n while x runs through V_n.

Proof. The equivalence of (i) and (ii): By (ii) of Theorem 2, a quadratic function has a nonlinearity larger than or equal to $2^{n-1} - 2^{\frac{1}{2}(n-1)}$ if and only if its linearity dimension is 1.

The equivalence of (ii) and (iii): Let $\alpha_1, \ldots, \alpha_{2^n-1}$ be the $2^n - 1$ nonzero vectors in V_n and $g_1, \ldots, g_{2^n-1}$ be the $2^n - 1$ nonzero linear combinations of $f_1, \ldots, f_n$. Similarly to the proof of Theorem 5, we construct a bipartite graph Γ with $\alpha_1, \ldots, \alpha_{2^n-1}$ on one side and $g_1, \ldots, g_{2^n-1}$ on the other side. A link exists between α_i and g_j if and only if α_i is a linear structure of g_j. Since each g_j is balanced, it must not be a bent function. By Lemma 5, each g_j has at least one nonzero linear structure. From the construction of Γ, we can see that each g_j has an edge associated with it. On the other hand, for any nonzero vector,

say α, $F(x) \oplus F(x \oplus \alpha)$ does not run through the vector zero, as $F(x)$ is a permutation on V_n. By Theorem 1, there exists a nonzero linear combination of the component functions of $F(x) \oplus F(x \oplus \alpha)$, say

$$\sum_{j=1}^{n} c_j[f_j(x) \oplus f_j(x \oplus \alpha)], \tag{5}$$

that is not balanced. Since f_j is quadratic, (5) is affine. Thus (5) must be a constant. Write $g_\alpha(x) = \sum_{j=1}^{n} c_j f_j(x)$. Then α is a nonzero linear structure of g_α. Thus each α has at least one edge associated with it. In summary, each g_j has at least one edge associated with it, and so does each α_j. As both sides of the bipartite graph have the same number of edges, (ii) and (iii) must stand in parallel.

The equivalence of (iii) and (iv): First we note that the differential uniformity of any permutation is at least 2. Let $s = n$ and $t = 1$. Then By Theorem 4, F is differential 2-uniform if and only if each nonzero vector in V_n is the linear structure of at most one nonzero linear combination of $f_1, \ldots, f_n$. In the proof of the equivalence of (ii) and (iii), it is has been shown that each nonzero vector in V_n is a linear structure of at least one nonzero linear combination of the component functions. Thus F is differential 2-uniform if and only if each nonzero vector in V_n is the linear structure of a unique nonzero linear combination of the component functions. □

In [15] Nyberg and Knudsen considered quadratic permutations on V_n that have the following property (P): *Every nonzero linear combination of the component functions, say g, can be expressed as $g(x) = xCx^T$, where $x = (x_1, \ldots, x_n)$, C is a nonsingular matrix over $GF(2)$ and the rank of $C \oplus C^T$ is $n-1$.* They proved that if a quadratic permutation on V_n satisfies the property (P) then it is a differentially 2-uniform permutation. From Theorem 6, we conclude that the property (P) is equivalent to (i), (ii), (iii) and (iv) in the theorem.

The following is another important corollary of Theorem 6.

Corollary 2 *There exists no differentially 2-uniform quadratic permutation on an even dimensional vector space.*

Proof. Let $F(x) = (f_1, \ldots, f_n)$ be a differentially 2-uniform quadratic permutation on V_n. By Theorem 6, each component function f_i has a unique nonzero linear structure. Hence the linearity dimension of f_i is 1, and the corollary follows immediately from Part (i) of Theorem 2. □

This gives a negative answer to an open problem regarding the existence of differentially 2-uniform quadratic permutations on an even dimensional vector space.

Now it is a right place to point out an error in [2]. Corollary 2 of [2] states that the permutation defined by a polynomial $P(x) = x^{2^\ell(2^k+1)}$ is a differentially 2-uniform quadratic permutation, where $x \in GF(2^n)$, ℓ, k and n are positive integers, and $\gcd(2^k + 1, 2^n - 1) = \gcd(k, n) = 1$. Beth and Ding claim that their corollary indicates the existence of differentially 2-uniform quadratic permutations on V_n, n even. This seemingly contradicts the non-existence result shown in our Corollary 2. However, one can see that when n is even, k must be odd in order for $\gcd(k, n) = 1$ to stand. On the other hand, if n is even and k is odd, then $\gcd(2^k + 1, 2^n - 1)$ has 3 as a factor. Thus $\gcd(2^k + 1, 2^n - 1) = \gcd(k, n) = 1$

can not stand for n even. In other words, Beth and Ding's corollary does not imply the existence of differentially 2-uniform quadratic permutations on V_n, n even.

The following result has been pointed out by these authors in [21]. It is included here, together with its proof, for the sake of completeness.

Theorem 7 *Let $F = (f_1, \ldots, f_n)$ $(n \geqq 3)$ be a differentially 2-uniform quadratic permutation. Then there exists a nonsingular matrix A of order n over $GF(2)$ such that $\Psi(x) = F(xA) = (f_1(xA), \ldots, f_n(xA)) = (\psi_1(x), \ldots, \psi_n(x))$ is also differentially 2-uniform, and each component function ψ_j satisfies the SAC.*

Proof. When $n \geqq 7$, it directly follows from Theorem 5. The proof described below applies to all $n \geqq 3$.

Let Φ denote the set of vectors γ such that $f_j \oplus f_j(x \oplus \gamma)$ is not balanced for some $1 \leqq j \leqq n$. By (ii) and (iii) of Theorem 6, we have $|\Phi| = n$. Since $|\Phi| < 2^{n-1}$ for all $n \geqq 3$, by Theorem 2 of [21], there exists a nonsingular matrix A of order n over $GF(2)$ that transforms F into a SAC-fulfilling S-box. □

4 Conclusion

We have proved that for quadratic S-boxes, there are close relationships among differential uniformity, linear structures, nonlinearity and the SAC. We have shown that by using our proof techniques, all differentially 2-uniform quadratic permutations can be treated in a unified fashion. In particular, general results regarding nonlinearity characteristics of these permutations are derived, regardless of the actual methods for constructing the permutations.

A future research direction is to extend the results to the more general case where component functions of an S-box can have an algebraic degree larger than 2. Another direction is to enlarge the scope of nonlinearity criteria examined so that it includes other cryptographic properties such as algebraic degree, propagation characteristics, and correlation immunity.

References

[1] C. M. Adams. On immunity against Biham and Shamir's "differential cryptanalysis". *Information Processing Letters*, 41:77–80, 1992.

[2] T. Beth and C. Ding. On permutations against differential cryptanalysis. In *Advances in Cryptology - EUROCRYPT'93*. Springer-Verlag, Berlin, Heidelberg, New York, 1993. to appear.

[3] E. Biham and A. Shamir. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, Vol. 4, No. 1:3–72, 1991.

[4] E. Biham and A. Shamir. *Differential Cryptanalysis of the Data Encryption Standard*. Springer-Verlag, New York, Berlin, Heidelberg, London, Paris, Tokyo, Hong Kong, Barcelona, Budapest, 1993.

[5] L. Brown, M. Kwan, J. Pieprzyk, and J. Seberry. Improving resistance to differential cryptanalysis and the redesign of LOKI. In *Advances in Cryptology - ASIACRYPT'91.* Springer-Verlag, Berlin, Heidelberg, New York, 1991. to appear.

[6] J. Detombe and S. Tavares. Constructing large cryptographically strong S-boxes. In *Advances in Cryptology - AUSCRYPT'92*, volume 718, Lecture Notes in Computer Science, pages 165–181. Springer-Verlag, Berlin, Heidelberg, New York, 1993.

[7] J. F. Dillon. A survey of bent functions. *The NSA Technical Journal*, pages 191–215, 1972. (unclassified).

[8] J.-H. Evertse. Linear structures in blockciphers. In *Advances in Cryptology - EUROCRYPT'87*, volume 304, Lecture Notes in Computer Science, pages 249–266. Springer-Verlag, Berlin, Heidelberg, New York, 1988.

[9] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes.* New York: North-Holland, 1977.

[10] M. Matsui. Linear cryptanalysis method for DES cipher. In *Advances in Cryptology - EUROCRYPT'93.* Springer-Verlag, Berlin, Heidelberg, New York, 1993. to appear.

[11] W. Meier and O. Staffelbach. Nonlinearity criteria for cryptographic functions. In *Advances in Cryptology - EUROCRYPT'89*, volume 434, Lecture Notes in Computer Science, pages 549–562. Springer-Verlag, Berlin, Heidelberg, New York, 1990.

[12] K. Nyberg. Perfect nonlinear S-boxes. In *Advances in Cryptology - EUROCRYPT'91*, volume 547, Lecture Notes in Computer Science, pages 378–386. Springer-Verlag, Berlin, Heidelberg, New York, 1991.

[13] K. Nyberg. On the construction of highly nonlinear permutations. In *Advances in Cryptology - EUROCRYPT'92.* Springer-Verlag, Berlin, Heidelberg, New York, 1992. to appear.

[14] K. Nyberg. Differentially uniform mappings for cryptography. In *Advances in Cryptology - EUROCRYPT'93.* Springer-Verlag, Berlin, Heidelberg, New York, 1993. to appear.

[15] K. Nyberg and L. R. Knudsen. Provable security against differential cryptanalysis. In *Advances in Cryptology - CRYPTO'92*, volume Lecture Notes in Computer Science. Springer-Verlag, Berlin, Heidelberg, New York, 1992. to appear.

[16] J. Pieprzyk. Bent permutations. In *Proceeding of the International Conference on Finite Fields, Coding Theory, and Advances in Communications and Computing*, Las Vegas, 1991.

[17] O. S. Rothaus. On "bent" functions. *Journal of Combinatorial Theory*, Ser. A, 20:300–305, 1976.

[18] J. Seberry, X. M. Zhang, and Y. Zheng. Nonlinearly balanced boolean functions and their propagation characteristics. In *Advances in Cryptology - CRYPTO'93.* Springer-Verlag, Berlin, Heidelberg, New York, 1993. to appear.

[19] J. Seberry, X. M. Zhang, and Y. Zheng. On constructions and nonlinearity of correlation immune functions. In *Advances in Cryptology - EUROCRYPT'93*. Springer-Verlag, Berlin, Heidelberg, New York, 1993. to appear.

[20] J. Seberry, X. M. Zhang, and Y. Zheng. Systematic generation of cryptographically robust s-boxes. In *Proceedings of the first ACM Conference on Computer and Commnications Security*, pages 172 – 182. The Association for Computing Machinery, New York, 1993.

[21] J. Seberry, X. M. Zhang, and Y. Zheng. Improving the strict avalanche characteristics of cryptographic functions. *Information Processing Letters*, 1994. (to appear).

[22] J. Seberry, X. M. Zhang, and Y. Zheng. Nonlinearity and propagation characteristics of balanced boolean functions. To appear in *Information and Computation*, 1994.

[23] A. F. Webster. Plaintext/ciphertext bit dependencies in cryptographic system. Master's Thesis, Department of Electrical Engineering, Queen's University, Ontario, Cannada, 1985.

[24] A. F. Webster and S. E. Tavares. On the designs of S-boxes. In *Advances in Cryptology - CRYPTO'85*, volume 219, Lecture Notes in Computer Science, pages 523–534. Springer-Verlag, Berlin, Heidelberg, New York, 1986.

Appendix

A Proof for Theorem 1

First we have

Lemma 7 *Let $L_i = (h_{i1}, \ldots, h_{i2^s})$ be the sequence of a linear function on V_s, where $i = 1, \ldots, 2^n$ $(n \geqq s)$. Set*

$$M = [L_1^T, \ldots, L_{2^n}^T].$$

If the rows of M are mutually orthogonal then each linear sequence of length 2^s appears as 2^{n-s} columns of M.

Proof. Let $\eta = (a_1, \ldots, a_{2^s})$ be a (1, -1) sequences of length 2^s. Since $\langle \eta, L_i \rangle = \sum_{p=1}^{2^s} a_p h_{ip}$, we have

$$\langle \eta, L_i \rangle^2 = 2^s + 2 \sum_{p<q} a_p a_q h_{ip} h_{iq}$$

and

$$\sum_{i=1}^{2^n} \langle \eta, L_i \rangle^2 = 2^{n+s} + 2 \sum_{i=1}^{2^n} \sum_{p<q} a_p a_q h_{ip} h_{iq} = 2^{n+s} + 2 \sum_{p<q} \sum_{i=1}^{2^n} a_p a_q h_{ip} h_{iq}.$$

Since rows of M are mutually orthogonal, we have $\sum_{j=1}^{2^n} h_{ip} h_{iq} = 0$ $(p \neq q)$ and hence

$$\sum_{j=1}^{2^n} \langle \eta, L_i \rangle^2 = 2^{n+s}. \tag{6}$$

Now suppose that L, an arbitrary linear sequence of length 2^s, appears as k columns of M. By noting

$$\langle L, L_i \rangle = \begin{cases} 2^s & \text{if } L = L_i \\ 0 & \text{otherwise} \end{cases}$$

we have

$$\sum_{j=1}^{2^n} \langle L, L_i \rangle^2 = k \cdot 2^{2s}. \tag{7}$$

Compare (6) and (7) we have

$$k \cdot 2^{2s} = 2^{n+s}$$

and hence $k = 2^{n-s}$. □

Note that (7) can be viewed as a generalization of Parseval's equation (Page 416, [9]). The following is the proof for Theorem 1.

Proof. (**for Theorem 1**) Suppose that F is a regular S-box, namely, $F(x)$ runs through all vectors in V_s, each precisely 2^{n-s} times, while x runs through V_n, where $x = (x_1, \ldots, x_n)$. Then the truth table of each component function f_i must contain an equal number of ones and zeros, i.e., f_i is balanced.

Now we show that any nonzero linear combination, $f(x) = \sum_{j=1}^{s} c_j f_j(x)$, of the s component functions is also balanced. Recall that for any nonsingular matrix A of order s, $(f_1(x), \ldots, f_s(x))$ is regular if and only if $(f_1(x), \ldots, f_s(x))A$ is (see Lemma 4).

Now suppose that the first column of A is $(c_1,\ldots,c_s)^T$. Let $G(x) = (g_1(x),\ldots,g_s(x)) = (f_1(x),\ldots,f_s(x))A$. Then G is also regular, and hence its first component function $g_1(x) = f(x) = \sum_{j=1}^{s} c_j f_j(x)$ is balanced. This proves one direction of the theorem.

We now prove the other direction. Suppose that all nonzero linear combinations of the component functions are balanced. Let

$$\xi_i = (c_{i1},\ldots,c_{i2^n})$$

be the truth table of f_i, $i = 1,\ldots,s$. ¿From the s truth tables, we construct 2^n linear functions on V_s as follows:

$$\varphi_j(y) = c_{1j}y_1 \oplus c_{2j}y_2 \oplus \cdots \oplus c_{sj}y_s \tag{8}$$

where $y = (y_1,\ldots,y_s)$ and $j = 1,\ldots 2^n$.

Let

$$\eta_j = (b_{j1},\ldots,b_{j2^s})$$

be the truth table of φ_j. Set

$$N = [\eta_1^T,\ldots,\eta_{2^n}^T].$$

Note that N is a $2^s \times 2^n$ matrix whose elements come from $GF(2)$.

N is constructed in such a way that its rows consist of precisely the 2^s different linear combinations of $\xi_1,\ldots,\xi_s$. To prove this is true, we take a close look at the rows of N. Let $\gamma_i = (b_{1i}, b_{2i},\ldots,b_{2^n i})$ be the ith row of N, $0 \leqq i \leqq 2^s - 1$. Since $b_{ji} = \varphi_j(\alpha_i)$, where α_i is the vector in V_s corresponding to the integer i, we have $\gamma_i = (\varphi_1(\alpha_i),\varphi_2(\alpha_i),\ldots,\varphi_{2^n}(\alpha_i))$. Write $\alpha_i = (a_{i1},\ldots a_{i2^s})$. Then

$$\begin{aligned}
\gamma_i &= (\sum_{j=1}^{s} c_{j1}a_{ij}, \sum_{j=1}^{s} c_{j2}a_{ij},\ldots,\sum_{j=1}^{s} c_{j2^n}a_{ij}) \\
&= \sum_{j=1}^{s} a_{ij}(c_{j1}, c_{j2},\ldots,c_{j2^n}) \\
&= \sum_{j=1}^{s} a_{ij}\xi_j.
\end{aligned}$$

This proves that γ_i, the ith row of N, is indeed a linear combination of $\xi_1,\ldots,\xi_s$. On the other hand, since any nonzero linear combination of $\xi_1,\ldots,\xi_s$ is balanced, $\xi_1,\ldots,\xi_s$ are linearly independent. Thus $\gamma_i \neq \gamma_j$ for any $i \neq j$. This proves our claim that the rows of N consist of precisely the 2^s different linear combinations of $\xi_1,\ldots,\xi_s$.

Now let M be an matrix obtained from N by substituting 0 with $+1$ and 1 with -1. Note that the sum of two different rows of N is a nonzero linear combination of $\xi_1,\ldots,\xi_s$ and hence balanced. This implies that the rows of M is mutually orthogonal. By Lemma 7 each linear sequence of length 2^s appears as 2^{n-s} columns of M. This in turn implies that the truth table of a linear function on V_s appears as 2^{n-s} columns of N, i.e. any linear function φ on V_s appears 2^{n-s} times in the set $\{\varphi_1,\ldots\varphi_{2^n}\}$, where φ_j is defined in (8). As there is a one to one correspondence between linear functions on V_s and vectors in V_s, we conclude that $F(x) = (f_1(x),\ldots,f_s(x))$ runs through all vectors in V_s, each 2^{n-s} times, while x runs through V_n. □